3 Kanter DS, Mikkola KM, Patel SR, et al. Thrombolytic therapy for pulmonary embolism: frequency of intracranial treatment of haemorrhage and associated risk factor. *Chest* 1997; **111:** 1241–45.
4 Foster PF, Moore LR, Sankary HN. Central venous catheterisation in patient with coagulopathy. *Arch Surg* 1992; **127:** 273–75.

Study of genes and environmental factors in complex diseases

Sir—David Clayton and Paul McKeigue (Oct 20, p 1356)[1] thoughtfully review epidemiological methods for studies of genes and environmental factors with ideas about trade-offs between case-control and cohort studies. We would like to augment this review from our perspective as cancer epidemiologists who have worked on these issues for some time.[2,3]

Case-control studies are cheaper and quicker from planning to completion than cohort studies. They allow more thorough ascertainment of disease and standard collection of biospecimens, such as frozen tissue at diagnosis.[2] To ensure that a study is "correctly designed"[1] and done is not always easy, given difficulties in case ascertainment, control selection, and participation.[3] In addition, biospecimens that are markers for exposure might be affected by treatment, limiting the usefulness of case-control studies.[2]

Self-report or proxy reports on exposure obtained after diagnosis in case-control studies can lead to differential misclassification with important consequences. Multiplicative interaction is attenuated rather than exaggerated by differential misclassification of exposure;[1,4] this is small comfort when we realise that this property does not hold for the exposure main effect, the joint effects, the effect of one factor in subgroups defined by the other, the effect of genotype adjusted for exposure, or for assessment of additive interaction.[5] Seriously misclassified exposure, whether differential or non-differential, undermines Clayton and McKeigue's goal of testing hypotheses about causal pathways amenable to intervention.

Much of the economic and ultimate public-health importance of cohort studies arises from their ability to study multiple endpoints in the same base population. Even if the cost is substantially higher than for case-control studies, the attendant gain in efficiency over time is substantial if the cohort is maintained.[3] Also, cohorts allow case-cohort or case-control studies to be nested within, providing an appropriate roster from which to select one or more controls per case, effectively eliminating the problems of control selection and participation in stand-alone case-control studies, which Clayton and McKeigue downplay by assuming that they are well designed. With control-to-case ratios as low as 4 or 5, the power of these efficient nested designs can approach that of the full cohort.

Of course, there remain many circumstances in which case-control studies are indicated, such as when the outcome is rare, information on a specific exposure is not collected in sufficient detail in cohort studies, or when biospecimens that cannot be readily obtained in cohort studies are needed. However, 1 000 000 people were enrolled in prospective cohort studies with blood-sample collections and questionnaire data on important chronic disease risk factors by 1999, and there may be more than 2 000 000 enrolled within 5 years.[3] These studies should provide important information on the environmental and genetic contributions, and their inter-relations, to common sources of mortality and morbidity, at least in adults in developed and in some developing countries.

The presence of these resources, particularly when collaboration is encouraged and analyses are coordinated, should allow investigators to focus case-control efforts where they can provide unique and not duplicative information.

*Sholom Wacholder, Montserrat Garcia-Closas, Nathaniel Rothman
Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive, Boulevard, m/s 7244, Bethesda MD 20892, USA
(e-mail: Wacholder@NIH.gov)

1 Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358:** 1356–60.
2 Caporaso N, Rothman N, Wacholder S. Case-control studies of common alleles and environmental factors. *J Natl Cancer Inst Monogr* 1999; **26:** 25–30.
3 Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. *J Natl Cancer Inst Monogr* 1999; **26:** 39–42.
4 Garcia-Closas M, Thompson WD, Robins JM. Differential misclassification and the assessment of gene-environment interactions in case-control studies. *Am J Epidemiol* 1998; **147:** 426–33.
5 Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. *Cancer Epidemiol Biomarkers Prev* 1999; **8:** 1043–50.

Sir—We agree with most of David Clayton and Paul McKeigue's points,[1] but differ on certain crucial issues.

The potential distortion that can arise from systematic bias in the retrospective assessment of exposure status in case-control studies is important. Clayton and McKeigue argue that the simple test for an interaction representing a departure from an otherwise multiplicative model may be robust to such biases, provided that errors are independent of genotype, but this latter point is a crucial assumption.[2] Furthermore, such biases could seriously distort other features of the joint effect of an environmental and a genetic determinant.

We believe the extent to which the targeting of interventions in accordance with genotype will ultimately prove useful is as yet unclear. The appropriate action will depend on the multifactorial nature of the disease in question, and on the severity of its consequences for individuals, families, and society; the costs, risks, and unrelated benefits associated with the intervention being considered; and the costs and risks associated with genetic and other screening to detect high-risk individuals. Targeted therapeutic intervention sometimes may provide maximum health benefits and keep costs to a minimum. At other times, the whole-population approach may be preferred,[3] irrespective of individual genotype.

A key issue is the contest between the benefits of prospective exposure assessment (before disease onset) embodied in a cohort design, and the benefits accruing from the greater efficiency of a case-control design. This brings us to the fundamental purpose of BioBank UK.

If the sole aim were to study several specific causal hypotheses (possibly interactions) over 10 years, Clayton and McKeigue's case would be strong. However, the BioBank UK initiative is really about setting up a foundation for various bioscience projects over the next 20–30 years. Many projects will be nested case-control studies that will benefit from the prospective (and potentially repeated) exposure assessment and the ability to undertake detailed additional assessment in cases and a limited number of controls.

We share Clayton and McKeigue's reservations about the interpretation of statistical interactions, particularly when the correct scale of analysis is unknown. However, the current emphasis is more on being able to describe the joint effects of causal determinants.

We share Clayton and McKeigue's belief that a large prospective cohort would be an inefficient approach for a 10-year initiative. However, we believe that BioBank UK's originators had a